

**UNITED STATES PATENT APPLICATION FOR:**  
**FLASH MEMORY CELL WITH BURIED FLOATING GATE AND METHOD FOR**  
**OPERATING SUCH A FLASH MEMORY CELL**

**INVENTORS:**

**PETER HAGEMEYER**  
**WOLFRAM LANGHEINRICH**

**ATTORNEY DOCKET NUMBER: INFN/0060**

**CERTIFICATION OF MAILING UNDER 37 C.F.R. 1.10**

I hereby certify that this New Application and the documents referred to as enclosed therein are being deposited with the United States Postal Service on March 24, 2004, in an envelope marked as "Express Mail United States Postal Service", Mailing Label No. EV416702912US addressed to: Commissioner for Patents, Mail Stop PATENT APPLICATION, P.O. Box 1450, Alexandria, VA 22313-1450

  
\_\_\_\_\_  
Signature

Gero G. McClellan  
\_\_\_\_\_  
Name

March 24, 2004  
\_\_\_\_\_  
Date of signature

## **FLASH MEMORY CELL WITH BURIED FLOATING GATE AND METHOD FOR OPERATING SUCH A FLASH MEMORY CELL**

### **CROSS-REFERENCE TO RELATED APPLICATIONS**

[0001] This application is a continuation of co-pending PCT patent application No. PCT/EP02/09920, filed September 5, 2002, which claims the benefit of German patent application serial number 101 46 978.0, filed September 24, 2001. Each of the aforementioned related patent applications is herein incorporated by reference in their entireties.

### **BACKGROUND OF THE INVENTION**

#### **Field of the Invention**

[0002] The present invention relates to a programmable read-only memory cell having a channel layer arranged between a selection gate and a floating gate.

#### **Description of the Related Art**

[0003] In contrast to dynamic random access memory cells (DRAMs), programmable read-only memory cells according to the principle of a flash memory can retain the stored information even without an external power supply.

[0004] Conventional flash memories generally comprise a field-effect transistor (FET) having an additional floating gate, which is formed between the selection gate (control gate) of the FET and a channel layer which connects the two source/drain regions of the FET to one another.

[0005] In this case, in the programming mode of the memory cell, a specific charge is applied to the floating gate, which is insulated from its surroundings. The conductivity of the channel layer and thus the switching state of the FET are subsequently determined. A distinction is made between "normally on" and "normally off" memory cells depending on whether the charged floating gate closes or opens the channel of the FET. In this case, it is particularly simple to read a flash memory cell, since only the conductivity of the channel is checked for this purpose.

[0006] Despite these advantages over volatile memories, flash memories are not used ubiquitously. In particular, the significantly slower programming and erasing times of this type of memory compared with the programming and erasing times of volatile memories inhibit the spread of flash memory cells.

[0007] Furthermore, constructive problems arise in the case of combined memories, where, e.g., DRAM memory cells are also fabricated on a chip in addition to the flash memory cells, on account of the different technology sequence of the two memory cell types.

[0008] U.S. Patent No. 6,052,311 entitled "Electrically Erasable Programmable Read only Flash Memory" and U.S. Patent No. 6,011,288 entitled "Flash Memory Cell with Vertical Channels and Source/Drain Bus Lines" disclose flash memory cells with a reduced lateral extent. Both memory cells in each case have a floating gate formed in a trench between the source and drain regions of the respective memory cell and a selection gate arranged above the floating gate. In this case, the channels run below or laterally with respect to the floating gate.

[0009] JP 59 154071 A discloses a read-only memory. Similar read-only memories are described in JP 61 078169 A, U.S. Patent Nos. 5,488,243, 6,252,275 and 6,248,626. U.S. Patent No. 5,598,367 discloses an EPROM which utilizes a trench capacitor structure.

### **SUMMARY OF THE INVENTION**

[0010] It is an object of the invention to provide a read-only memory cell, an arrangement having such read-only memory cells, and a method for operating such a read-only memory cell which are distinguished by a high storage density and also by fast write and erase operations.

[0011] One object of the invention is to provide a flash memory cell which enables a higher storage density and a faster writing and erasing operation. Furthermore, it is an object of the invention to provide methods for operating such a flash memory cell.

[0012] According to one embodiment of the invention, the flash memory cell has a channel layer which is arranged between the floating gate and the selection gate and connects the source and drain electrodes to one another.

[0013] In this case, the floating gate arranged below the selection gate is arranged at least partly in a trench formed in the substrate. Vertically extending the trench into the substrate makes it possible to minimize the diameter of the floating gate and thus also the effective chip area of the memory cell.

[0014] In accordance with a further advantageous refinement of the invention, the memory cell has two separate oxide layers for the writing/erasing and reading operations. As a result of this, each of the two oxide layers and thus also the writing/erasing and reading operations connected with the respective oxide layer can be optimized separately, resulting not only in an improved tunnel oxide layer but also, in particular, shorter writing and erasing times becoming possible.

[0015] In accordance with a further advantageous refinement of the invention, the channel layer is formed as an epitaxial layer. This makes it possible to configure the channel layer to be so thin that a maximum control effect of selection gate and floating gate is achieved.

[0016] In accordance with a further refinement of the invention, the buried floating gate forms the inner electrode, a first diffusion region forms the outer electrode, and an insulator layer formed between the floating gate and the first diffusion region forms the dielectric of a trench capacitor extending into the substrate. Since the trench capacitor is formed in accordance with a trench capacitor of a DRAM memory cell, process steps can be saved during the fabrication of combined applications, where flash and DRAM memory cells are fabricated together on a semiconductor wafer. Furthermore, the constructive problems that are customary in the case of conventional flash memory cells are obviated in the case of these combined applications on account of the adapted dimensions of the two memory cell types.

[0017] On account of the construction of the flash memory cell according to the invention, in which the floating gate forms the inner electrode of a trench capacitor

and the floating gate is charged and discharged capacitively via a first diffusion region forming the outer electrode of the trench capacitor, the coupling area between the floating gate and the first diffusion region turns out to be particularly large. As a result of this, the floating gate can be capacitively charged and discharged particularly effectively.

[0018] In accordance with a further advantageous refinement of the invention, the first diffusion regions of adjacent memory cells of a series of the arrangement that is perpendicular to the word line direction overlap among one another. This gives rise to a second bit line along the row of memory cells, via which bit line each memory cell can be programmed or erased.

#### **BRIEF DESCRIPTION OF THE DRAWINGS**

[0019] So that the manner in which the above recited features of the present invention can be understood in detail, a more particular description of the invention, briefly summarized above, may be had by reference to embodiments, some of which are illustrated in the appended drawings. It is to be noted, however, that the appended drawings illustrate only typical embodiments of this invention and are therefore not to be considered limiting of its scope, for the invention may admit to other equally effective embodiments.

[0020] Figure 1 shows a cross section through a flash memory cell according to one embodiment of the invention with a buried floating gate;

[0021] Figures 2A to 2C show the method of operation of the flash memory cell according to one embodiment of the invention from Figure 1 during a writing operation, an erasing operation and a reading operation; and

[0022] Figure 3 shows a matrix-type arrangement of flash memory cells according to one embodiment of the invention with second bit lines formed by overlapping of the first diffusion regions.

## **DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT**

[0023] Figure 1 illustrates the construction of a flash memory cell MC according to the invention. The memory cell MC has a floating gate FG, buried within a substrate 10 and a field-effect transistor formed above the buried floating gate FG. The illustrated embodiment of the invention shows a "normally on" memory cell, the field-effect transistor being activated in the case of an uncharged floating gate FG.

[0024] In order to reduce the chip area, the floating gate FG is accommodated completely in a trench TR formed within the substrate 10, and at the same time forms the inner electrode of a trench capacitor 20.

[0025] A thin insulator layer 21 is formed within the trench TR. The insulator layer 21 completely covers the bottom and the sidewalls of the trench TR with a uniform layer thickness and reaches as far as the substrate surface. The insulator layer 21, which may be formed as an ONO layer (oxide-nitride-oxide), serves as a dielectric of the trench capacitor 20 and insulates the floating gate FG from a first diffusion region 22 forming the outer electrode of the trench capacitor 20.

[0026] In the exemplary embodiment illustrated, the first diffusion region 22 has an n-type doping and serves for capacitively charging and discharging the floating gate FG. In order to achieve a maximum coupling capacitance between the floating gate FG and the first diffusion region 22, the trench TR is completely surrounded by the first diffusion region 22 except for its topmost region. In this case, the first diffusion region 22 is formed as a well within the substrate 10 and extends from a level below the trench TR as far as a level just below the substrate surface.

[0027] As revealed by Figure 3, the first diffusion regions 22 of a row of a matrix-type arrangement of flash memory cells MC overlap one another and form a second bit line BL2 for writing to and erasing the flash memory cell MC.

[0028] A second diffusion region 23 is provided outside the first diffusion region 22, and extends from the substrate surface to a point below the first diffusion region 22 and laterally to a point beyond the flash memory cell MC. In this case, the second diffusion region 23 is illustrated in Figure 1 as a well comprising only a single

memory cell MC. As indicated in Figures 2A to 2C, the second diffusion region 23 may also extend to further memory cells MC of a matrix-type arrangement. In this case, the second diffusion region 23 is formed completely within a third diffusion region 24 formed as a well or areally in the substrate 10. In this case, the second diffusion region 23 has a p-type doping and the third diffusion region 24 has an n-type doping. The specific arrangement of the diffusion regions 22, 23, 24 forms a "triple well" arrangement, the first diffusion region 22 and the third diffusion region 24 being electrically insulated from one another, independently of their respective charge states, on account of depletion layers which form at the pn junctions between the diffusion regions 22, 23, 24. A similar arrangement is formed by the n-doped source/drain electrodes S, D with the first and second diffusion regions 22, 23. In this case, the first diffusion region 22 is electrically insulated from the source/drain electrodes S, D likewise on account of depletion layers which form at the pn junctions between the diffusion regions 22, 23 and the source/drain electrodes S, D.

[0029] A thin insulator layer TOX is formed above the floating gate FG at the level of the substrate surface, and completely covers the floating gate FG. The insulator layer TOX forms the tunnel oxide of the flash memory cell MC, which charges and discharges the floating gate FG, forming the inner electrode of the trench capacitor 20, during writing and erasing operations. The thickness of the tunnel oxide layer TOX is chosen such that, on the one hand, the charge situated on the floating gate FG is insulated sufficiently well from a conductive channel layer EPI of the field-effect transistor (FET) and, on the other hand, a sufficiently high tunneling current is ensured during writing and erasing operations of the memory cell MC.

[0030] On the substrate surface, a field-effect transistor (FET) is formed above the buried floating gate FG. The source electrode S of the transistor is arranged on one side of the memory trench TR, and the drain electrode D of the transistor is arranged on the other side of said memory trench TR. A channel layer EPI extends between the source and drain electrodes S, D and electrically connects the two electrodes S, D to one another. In this case, the channel layer EPI may cover the entire tunnel layer TOX, the upper partial regions of the insulator layer 21 formed as an ONO layer, and also partial regions of the substrate surface adjoining the trench TR. In

this case, the channel layer EPI may be composed of epitaxial silicon and may have an n-type doping.

[0031] A selection gate CG is formed above the channel layer EPI. The selection gate CG and the channel layer EPI are isolated from one another by a gate oxide layer GOX situated in-between. The gate oxide layer GOX formed as a thin insulator layer covers the entire channel layer EPI and also partial regions of the two source/drain electrodes S, D. A word line WL is formed above the selection gate CG and connects to one another the memory cells MC of a column of the matrix-type arrangement of memory cells MC that is illustrated in Figure 3. In this case, the word line WL serves for the addressing of the memory cells MC in the y-direction.

[0032] The substrate surface is covered with a further insulator layer 11, in which the entire FET structure is also embedded. For the contact-connection of the source/drain electrodes S, D, a first and a second contact 30, 31 are formed in the insulator layer 11. The second contact 31 may be connected to a first bit line BL1. In this case, the first bit line BL1 (not illustrated) may run orthogonally with respect to the word lines WL of the matrix-type arrangement of memory cells MC as illustrated in Figure 3, and in this case serves for the addressing in the x-direction.

[0033] Figure 2A diagrammatically shows the writing operation of a flash memory cell MC analogous to that illustrated in Figure 1. During a writing operation, the floating gate FG is charged negatively. For this purpose, electrons migrate from the channel layer EPI into the floating gate FG and, in the process, tunnel through the tunnel oxide layer TOX under a high electric field, generated by the pulling voltage  $U_{\text{program}}$  formed between the channel layer EPI and the first diffusion region 22.

[0034] In order to generate the required pulling voltage  $U_{\text{program}}$ , the source/drain electrodes S, D may be put jointly at a negative potential  $-\Phi_{\text{program}}$ . By the application of a positive potential  $\Phi_{\text{ON}}$  to the selection gate CG, a conductive n-channel 32 is produced within the channel layer EPI, as a result of which the channel layer EPI, which forms one of the two tunnel electrodes, is likewise brought to the source/drain potential  $-\Phi_{\text{program}}$ . The second tunnel electrode is formed by the first diffusion region 22. In order to generate the pulling voltage  $U_{\text{program}}$ , the first diffusion region 22 is put

at a positive potential  $+\Phi_{\text{program}}$  by a second bit line BL2. In this case, the second bit line BL2 is formed by overlap regions 22a, as illustrated in Figure 3, of the first diffusion regions 22 of directly adjacent memory cells MC of a row of the arrangement that is perpendicular to the word line direction.

[0035] As a result of the large coupling area of the trench capacitor, the capacitive interaction between the first diffusion region 22 and the floating gate FG is so large in the floating gate FG that such a high positive potential is induced in the floating gate FG that electrons can tunnel through the tunnel oxide layer TOX.

[0036] The tunneling electrons charge the floating gate FG negatively. Since the floating gate FG is electrically insulated from its surroundings, the electrons remain within the floating gate FG even after the supply voltage has been switched off. The electric field strengths occurring between the channel layer EPI and the floating gate FG in reading operation of the memory cell MC generally do not suffice to discharge the floating gate FG again via the tunnel oxide layer TOX.

[0037] Therefore, the information unit (bit) written in the memory cell MC is ideally preserved indefinitely or until the intentional discharge of the memory cell.

[0038] Figure 2B diagrammatically shows the erasing operation of the flash memory cell MC illustrated in Figure 2A. In order to erase the information unit of the memory cell, the trench capacitor 20 is discharged again. In this case, the electrons that have tunneled from the floating gate FG tunnel via the tunnel oxide layer TOX into the channel layer EPI. In this case, the electrons are pulled by a high pulling voltage  $U_{\text{erase}}$  formed between the first diffusion region 22 and the channel layer EPI. For this purpose, the source and drain electrodes S, D are jointly put at a positive electrical potential  $+\Phi_{\text{erase}}$ . Analogously to the writing operation illustrated in Figure 2A, during the erasing operation a conductive n-channel 32 is produced in the channel layer EPI by the application of a positive electrical potential  $\Phi_{\text{ON}}$  to the selection gate CG. As a result of this, the channel layer EPI, which forms one tunnel electrode, likewise acquires the positive electrical potential  $+\Phi_{\text{erase}}$ . By contrast, the diffusion region 22 forming the second tunnel electrode is put at a negative potential  $-\Phi_{\text{erase}}$  via the

second bit line BL2, which is illustrated in Figure 3. On account of the high capacitive interaction between the first diffusion region 22 and the floating gate FG, a sufficiently high negative potential is induced in the upper region of the floating gate FG, so that electrons tunnel through the tunnel oxide layer EPI. As a result of this, the floating gate FG is completely discharged again and the memory cell MC is brought to the initial state "normally on" again.

[0039] Figure 2 diagrammatically shows the reading operation of the flash memory cell MC. During the reading of the information stored in the memory cell MC, the conductivity of the channel layer EPI between the selection and floating gates CG, FG is evaluated. The memory cell MC is assigned one of the two logic data units "1" or "0", depending on the charge state of the floating gate FG and the resultant conductance of the channel 32. In the case of the "normally on" memory cell MC illustrated here, the channel 32 is blocked when the trench capacitor 20 is charged, and open when the trench capacitor 20 is discharged.

[0040] In order to read the flash memory cell MC, a read voltage  $U_{\text{read}}$  is generated between the source and drain electrodes S, D. The source electrode S may be put at the ground potential  $\Phi_{\text{ground}}$ , and the drain electrode D may be put at a positive potential  $+\Phi_{\text{read}}$ . In this case, the selection gate CG at the first diffusion region 22 may acquire the same electrical potential  $+\Phi_{\text{read}}$  as the drain electrode D.

[0041] On account of the influence field generated by the electrical potential  $+\Phi_{\text{read}}$  of the selection gate CG, the channel 32 is open in the case of an uncharged floating gate FG. In this case, a detectable current flow arises in the channel layer EPI on account of the read voltage  $U_{\text{read}}$  present between the source and drain electrodes S, D.

[0042] By contrast, if the floating gate FG has a negative charge, then the channel 32 within the channel layer EPI is pinched off by the influence field of the negative charge. The conductivity of the channel layer EPI is thereby reduced. The charge state of the memory cell MC is then detected on the basis of a significantly reduced or totally stopped current flow between the source and drain electrodes S, D.

[0043] The conductivity of the channel layer EPI, which corresponds to the charge state of the memory cell MC, is determined, in both cases, by a conventional evaluation circuit which, in the simplest case, checks whether a current flows between the source and drain electrodes S, D. If this is the case, then the memory cell MC is assigned an information unit "1" or "0", depending on the memory cell concept. Otherwise, the memory cell MC is assigned the respective complementary information unit.

[0044] Figure 3 shows a plan view of a matrix-type arrangement of flash memory cells MC. In this case, the memory cells MC are arranged in in each case four columns and rows running perpendicularly to one another. A trench isolation STI, which electrically isolates the memory cells MC of a column from one another, is formed between two directly adjacent rows of the arrangement. Each of the memory cells MC of the arrangement is formed analogously to the flash memory cell MC illustrated in Figure 1 and has in each case a floating gate FG formed in a trench TR of the substrate 10. The floating gate FG is electrically insulated from a first diffusion region 22 by an insulator layer 21. A channel layer EPI is in each case arranged above the floating gate FG, the floating gate FG being isolated from the channel layer EPI by a thin tunnel oxide layer TOX. Each channel layer EPI may be formed as an epitaxial layer and in each case connects two source/gate electrodes S, G to one another, which are arranged on both sides of the channel layer EPI. In this case, each of the source/drain electrodes S, D is assigned to two directly adjacent memory cells MC of a row of the arrangement that runs perpendicularly to the word line direction. Above the channel layer EPI, each memory cell MC has a selection gate CG, which is isolated from the channel layer EPI by a thin gate oxide layer GOX.

[0045] The memory cells MC within the matrix-type arrangement are addressed in the y-direction in each case by a word line WL. In this case, the word line WL makes contact with all the selection gates CG of the memory cells MC of a column of the arrangement.

[0046] First bit lines BL1 (not illustrated in Figure 3) are arranged orthogonally with respect to the word lines WL and in each case make contact with the source/drain electrodes S, D of the memory cells MC of a row of the arrangement.

[0047] The first diffusion regions 22 of each memory cell MC in each case have an overlap region 22a with the first diffusion regions 22 of the two directly adjacent memory cells MC of the respective row of the arrangement that runs perpendicularly to the word line direction. The electrically conductive connection produced in this way forms a second bit line BL2, via which information is written to the memory cell MC or erased from the memory cell MC. To that end, as revealed by the description of Figures 2A and 2B, the first diffusion region 22 acquires a positive or negative electrical potential  $+\Phi_{\text{program}}$ ,  $-\Phi_{\text{erase}}$ , respectively, via the second bit line BL2 assigned to the respective memory cell MC.

[0048] In order to carry out a reading operation, each memory cell MC of the matrix-type arrangement can be addressed individually with the aid of the word lines WL and the first bit lines BL1. The respective second bit line BL2 is additionally necessary for carrying out the writing or erasing operation of the respective memory cell MC.

[0049] While the foregoing is directed to embodiments of the present invention, other and further embodiments of the invention may be devised without departing from the basic scope thereof, and the scope thereof is determined by the claims that follow.